[Patent Document]

1. Japanese Patent Laid Open No. 2003-303055

Disk storage system having disk arrays connected with disk adaptors through switches

Hitachi, Ltd.

Inventor(s):Tanaka, Katsuya ;Fujimoto, Kazuhisa

Application No. 10/212882, Filed 20020807, A1 Published 20031009

Abstract:

A disk storage system has high throughput between a disk adapter of a disk controller and a disk array. The disk adapter of the disk controller is connected to the disk array through switches. Data on a channel between the switch and a RAID group is multiplexed in the switch to be transferred onto a channel between the switch and the disk adapter and data on the channel between the switch and the disk adapter is demultiplexed in the switch to be transferred onto the channel between the switch and the RAID group. A data transfer rate on the channel between the disk adapter and the switch is made higher than that on the channel.

US.Class: 711114 711154

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開2003－303055
(P2003－303055A)
(43)公開日　平成15年10月24日(2003.10.24)

(51)Int.Cl.'	識別記号	ＦＩ	テーマコード'(参考)
Ｇ０６Ｆ　3/06	３０１	Ｇ０６Ｆ　3/06	３０１Ｍ　５Ｂ０６５
			３０１Ｂ
	３０２		３０２Ａ
	５４０		５４０

審査請求　未請求　請求項の数９　ＯＬ　（全 14 頁）

(21)出願番号　特願2002－106262(P2002－106262)

(22)出願日　平成14年４月９日(2002.4.9)

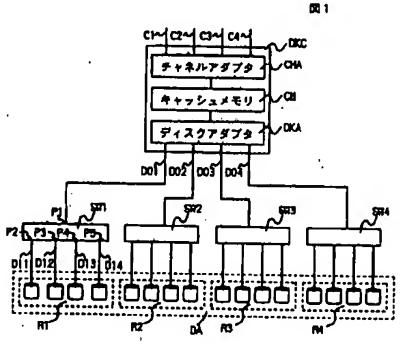
(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地
(72)発明者　田中　勝也
東京都国分寺市東恋ケ窪一丁目280番地
株式会社日立製作所中央研究所内
(72)発明者　藤本　和久
東京都国分寺市東恋ケ窪一丁目280番地
株式会社日立製作所中央研究所内
(74)代理人　100099298
弁理士　伊藤　悠　（外１名）

Ｆターム(参考)　5B065 BA01 CA04 CA07 CA12 CA15
CA30 CC08 CE11 CH01

(54)【発明の名称】　ディスクアダプタとディスクアレイをスイッチを介して接続したディスク装置

(57)【要約】
【課題】　ディスクコントローラのディスクアダプタとディスクアレイ間のスループットが高いディスク装置を提供することにある。
【解決手段】　ディスクコントローラ(DKC)のディスクアダプタ(DKA)とディスクアレイ(DA)をスイッチ(SW1,SW2,SW3,SW4)を介して接続する。スイッチ(SW1)とＲＡＩＤグループ(R1)間のチャネル(D11,D12,D13,D14)上のデータをスイッチ(SW1)において多重化してスイッチ(SW1)とディスクアダプタ(DKA)間のチャネル(D01)に転送し、スイッチ(SW1)とディスクアダプタ(DKA)間のチャネル(D01)上のデータをスイッチ(SW1)において逆多重化してスイッチ(SW1)とＲＡＩＤグループ(R1)間のチャネル(D11,D12,D13,D14)に転送する。ディスクアダプタ(DKA)とスイッチ(SW1)間のチャネル(D01)上のデータ転送速度を、チャネル(D11,D12,D13,D14)のデータ転送速度より高くする。

図1



【特許請求の範囲】
【請求項１】　ディスクコントローラとディスクアレイからなり、前記ディスクコントローラはチャネルアダプタとキャッシュメモリとディスクアダプタを有するディスク装置において、
前記ディスクアダプタと前記ディスクアレイを、バッファメモリを有するスイッチを介して接続し、
前記スイッチは、前記ディスクアダプタが接続されたポートと前記ディスクアレイを構成するディスクドライブが接続された各ポートとの間でのポート間の接続の切り換えを、入力されたフレーム毎に、該フレーム内の送信先情報にしたがって行うことを特徴とするディスク装置。
【請求項２】　ディスクコントローラと複数のディスクアレイからなり、前記ディスクコントローラはチャネルアダプタとキャッシュメモリとディスクアダプタを有するディスク装置において、
前記ディスクアレイはループ状に接続した複数のディスクドライブからなり、前記ディスクアダプタと前記複数のディスクアレイとをバッファメモリを有するスイッチを介して接続し、
前記ディスクアダプタと前記スイッチ間のチャネル当りデータ転送速度を、前記スイッチと前記複数のディスクアレイ間のチャネル当りデータ転送速度より高く設定し、
前記スイッチは、前記ディスクアダプタが接続されたポートと前記複数のディスクアレイが接続された各ポートとの間でのポート間の接続の切り換えを、入力されたフレーム毎に、該フレーム内の送信先情報にしたがって行うことを特徴とするディスク装置。
【請求項３】　ディスクコントローラとディスクアレイからなり、前記ディスクコントローラはチャネルアダプタとキャッシュメモリとディスクアダプタを有するディスク装置において、
前記ディスクアダプタと前記ディスクアレイを、バッファメモリを有するスイッチを介して接続し、
同一のスイッチに接続したディスクドライブの組み合わせでＲＡＩＤグループを構成し、
前記ディスクアダプタと前記スイッチ間のチャネル当りデータ転送速度を、前記スイッチと前記ディスクアレイ間のチャネル当りデータ転送速度より高く設定し、
前記スイッチは、前記ディスクアダプタが接続されたポートと前記ＲＡＩＤグループを構成するディスクドライブが接続された各ポートとの間でのポート間の接続の切り換えを、入力されたフレーム毎に、該フレーム内の送信先情報にしたがって行うことを特徴とするディスク装置。
【請求項４】　第１のディスクコントローラと第２のディスクコントローラと複数のディスクアレイからなり、第１のディスクコントローラは第１のチャネルアダプタと第１のキャッシュメモリと第１のディスクアダプタを有し、
第２のディスクコントローラは第２のチャネルアダプタと第２のキャッシュメモリと第２のディスクアダプタを有するディスク装置において、
第１のディスクアダプタと前記複数のディスクアレイとをバッファメモリを有する第１のスイッチを介して接続し、且つ第２のディスクアダプタと前記複数のディスクアレイとをバッファメモリを有する第２のスイッチを介して接続し、さらに第１のスイッチと第２のディスクアダプタを接続し、第２のスイッチと第１のディスクアダプタを接続し、
第１のディスクアダプタと第１のスイッチ間、および第２のディスクアダプタと第１のスイッチ間のチャネル当りデータ転送速度を第１のスイッチと前記複数のディスクアレイ間のチャネル当りデータ転送速度より高く設定し、
第２のディスクアダプタと第２のスイッチ間、および第１のディスクアダプタと第２のスイッチ間のチャネル当りデータ転送速度を第２のスイッチと前記複数のディスクアレイ間のチャネル当りデータ転送速度より高く設定し、
第１のスイッチは、第１のディスクアダプタまたは第２のディスクアダプタが接続されたポートと前記複数のディスクアレイが接続された各ポートとの間でのポート間の接続の切り換えを、入力されたフレーム毎に、該フレーム内の送信先情報にしたがって行い、
第２のスイッチは、第１のディスクアダプタまたは第２のディスクアダプタが接続されたポートと前記複数のディスクアレイが接続された各ポートとの間でのポート間の接続の切り換えを、入力されたフレーム毎に、該フレーム内の送信先情報にしたがって行うことを特徴とするディスク装置。
【請求項５】　第１のディスクコントローラと第２のディスクコントローラと複数のディスクアレイからなり、第１のディスクコントローラは第１のチャネルアダプタと第１のキャッシュメモリと第１のディスクアダプタを有し、
第２のディスクコントローラは第２のチャネルアダプタと第２のキャッシュメモリと第２のディスクアダプタを有するディスク装置において、
第１のディスクアダプタと前記複数のディスクアレイとをバッファメモリを有する第１のスイッチを介して接続し、且つ第２のディスクアダプタと前記複数のディスクアレイとをバッファメモリを有する第２のスイッチを介して接続し、さらに第１のスイッチと第２のディスクアダプタを接続し、第２のスイッチと第１のディスクアダプタを接続し、
第１のディスクアダプタと第１のスイッチ間、および第２のディスクアダプタと第１のスイッチ間のチャネル当

りデータ転送速度を第１のスイッチと前記複数のディスクアレイ間のチャネル当りデータ転送速度より高く設定し、

第２のディスクアダプタと第２のスイッチ間、および第１のディスクアダプタと第２のスイッチ間のチャネル当りデータ転送速度を第２のスイッチと前記複数のディスクアレイ間のチャネル当りデータ転送速度より高く設定し、

第１のスイッチと第２のスイッチを、第１のディスクアダプタと第２のスイッチ間を接続したチャネルと同等のデータ転送速度を有するチャネルと、第２のディスクアダプタと第１のスイッチ間を接続したチャネルと同等のデータ転送速度を有するチャネルと、を介して接続し、

第１のスイッチは、第１のディスクアダプタまたは第２のディスクアダプタまたは第２のスイッチが接続されたポートと前記複数のディスクアレイが接続された各ポートとの間でのポート間の接続の切り換えを、入力されたフレーム毎に、該フレーム内の送信先情報にしたがって行い、

第２のスイッチは、第１のディスクアダプタまたは第２のディスクアダプタまたは第１のスイッチが接続されたポートと前記複数のディスクアレイが接続された各ポートとの間でのポート間の接続の切り換えを、入力されたフレーム毎に、該フレーム内の送信先情報にしたがって行うことを特徴とするディスク装置。

【請求項６】　請求項１乃至請求項５のいずれかの請求項記載のディスク装置において、

前記ディスクアレイからのデータ読み出し時には、前記ディスクアレイから前記スイッチに転送されるデータを前記スイッチにおいて多重化して前記ディスクアダプタに転送し、

前記ディスクアレイへのデータ書き込み時には、前記ディスクアダプタから前記スイッチに転送されるデータを前記スイッチにおいて逆多重化して前記ディスクアレイに転送することを特徴とするディスク装置。

【請求項７】　請求項１乃至請求項５のいずれかの請求項記載のディスク装置において、

ディスクアダプタからディスクアレイへのデータ書き込み時に、前記ディスクアダプタは、前記ポート間の接続の切り替えが周期的に行われるように、送出するフレームに送信先情報を設定し、

ディスクアレイからディスクアダプタへのデータ読み出し時に、前記スイッチは、ラウンドロビン方式により前記ポート間の接続を切り替えることを特徴とするディスク装置。

【請求項８】　請求項７記載のディスク装置において、周期的に切り　えるポート数を、ディスクアダプタとスイッチ間のチャネル当りデータ転送速度の、スイッチとディスクアレイ間のチャネル当りデータ転送速度に対する比、と同程度に設定することを特徴とするディスク装置。

【請求項９】　請求項１乃至請求項５のいずれかの請求項記載のディスク装置において、

前記ディスクアダプタと前記スイッチ間を光ファイバケーブルで接続し、前記スイッチと前記ディスクアレイ間をメタルケーブルで接続することを特徴とするディスク装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、コンピュータシステムにおける２次記憶装置に関し、特に入出力データ転送性能が高いディスク装置に関する。

【０００２】

【従来の技術】現在のコンピュータシステムにおいては、ＣＰＵ（中央処理装置）が必要とするデータは２次記憶装置に保存され、ＣＰＵなどが必要とするときに応じて２次記憶装置に対してデータの書き込みおよび読み出しを行う。この２次記憶装置としては、一般に不揮発な記憶媒体が使用され、代表的なものとして磁気ディスク装置や、光ディスクなどのディスク装置がある。近年高度情報化に伴い、コンピュータシステムにおいて、この種の２次記憶装置の高性能化が要求されている。

【０００３】図９に、従来のディスク装置のブロック図を示す。図９において、ディスク装置はディスクコントローラＤＫＣとディスクアレイＤＡで構成される。ディスクコントローラＤＫＣは、上位側ＣＰＵ（図示せず）とディスク装置を接続するチャネルアダプタＣＨＡと、ディスクアレイＤＡに対して読み書きするデータを一時保存するキャッシュメモリＣＭと、ディスクコントローラＤＫＣとディスクアレイＤＡを接続するディスクアダプタＤＫＡからなる。チャネルアダプタＣＨＡとキャッシュメモリＣＭとディスクアダプタＤＫＡは、バスまたはスイッチで相互接続されている。チャネルアダプタＣＨＡはＣ１、Ｃ２、Ｃ３、Ｃ４の４本のチャネルでＣＰＵと接続している。ディスクアダプタＤＫＡはＤ１、Ｄ２、Ｄ３、Ｄ４の４本のチャネルでディスクアレイと接続している。ここでディスクアレイＤＡはディスクグループＲ１、Ｒ２、Ｒ３、Ｒ４からなる。ディスクアレイＤＡにおいてＲＡＩＤシステムを構築する場合は、Ｒ１、Ｒ２、Ｒ３、Ｒ４がそれぞれＲＡＩＤグループを構成する。

【０００４】チャネルＣ１、Ｃ２、Ｃ３、Ｃ４から入力された書き込みデータは、キャッシュメモリＣＭに該データを書き込むと同時に、該データをブロックサイズ単位に分割し、チャネルＤ１、Ｄ２、Ｄ３、Ｄ４の内３チャネルにはブロック単位に分割されたデータを、残りの１チャネルは前記分割データから計算したパリティを、ディスクアダプタＤＫＡからディスクアレイＤＡへ送信する。データ読み出し時は、先ずキャッシュメモリＣＭ内に該当データの有無を調べる。有る場合は、キャッシュメモリＣＭからチャネルアダプタＣＨＡを介してキャッシュメモリ内読み出しデータをＣＰＵへ送信する。キャッシュメモリＣＭ内に無い場合にディスクアダプタＤＫＡは、Ｄ１、Ｄ２、Ｄ３、Ｄ４を介してディスクアレイＤＡからブロック単位に分割されたデータを読み出し、チャネルアダプタＣＨＡを介して読み出しデータをＣＰＵへ送信する。　この種の従来技術を第１の従来技術と呼ぶ。第１の従来技術として関連するディスク装置は、例えば、日経ＢＰ社刊の「日経コンピュータ別冊メインフレーム’９８」（１９９８年）第１４４頁から第１５３頁に記載されているディスク装置がある。

【０００５】ディスクアダプタとディスクアレイを、スイッチを介して接続したディスク装置が、特開平５－１７３７２２号の「マルチチャンネルデータおよびパリティの交換デバイス」に開示されている。以下、該公報に記載の従来の技術を第２の従来技術と呼ぶ。第２の従来の技術によれば、ディスクアレイに関連したバス本数とディスクアダプタに関連したバス本数とを独立に設定できる。ディスクアダプタとディスクアレイを、バッファ制御ブロックを介して接続したディスク装置が、特開平６－１９６２７号の「回転形記憶装置」に開示されている。以下、該公報に記載の従来の技術を第３の従来技術と呼ぶ。第３の従来技術によれば、ディスクアダプタとディスクアレイ間のデータ転送速度を任意に設定でき、ディスクの回転待ちの影響を低減できる。

【０００６】

【発明が解決しようとする課題】ネットワーク技術の進歩に伴い、１チャネル当りのデータ転送速度は年々増加している。例えばディスク装置に使われるファイバチャネルでは、現状でチャネル当りのデータ転送速度が１Ｇｂｐｓから２Ｇｂｐｓであるが、近い将来４Ｇｂｐｓから１０Ｇｂｐｓへ高速化されることが予定されている。ＣＰＵとチャネルアダプタ間（以下フロントエンドと呼ぶ）のスループットはこの高速化に従うことが予想される。ところが、ディスクアダプタとディスクアレイ間（以下バックエンドと呼ぶ）のスループットは以下の理由により、フロントエンドほど高速化されないと予想される。第１の理由は、ディスクドライブは機械部品を含むので、電子、光素子のみ高速化を行えば良いフロントエンドに比べ高速化が難しいこと、である。第２の理由は、たとえディスクドライブが高速化したとしても、全てのディスクドライブ毎に高速インターフェイスを搭載するのは、多数のディスクドライブを有するディスク装置の高コスト化を招くことである。

【０００７】第１の従来技術では、チャネルアダプタのチャネル当りのデータ転送速度を向上させても、フロントエンドとバックエンドのスループット乖離により、ディスク装置の性能が向上しないという問題があった。また、バックエンドのスループット向上のために低速ポートを多数、ディスクアダプタに設けることも考えられるが、ディスクアダプタのポート数増加は制御を複雑とする。第２の従来技術では、ディスクアダプタとディスクアレイとの間にスイッチを適用することによりディスク増設ポート数を増加させることができるが、チャネル当りのデータ転送速度はディスクアレイのデータ転送速度に制限されるので、ディスクアダプタとディスクアレイ間のスループットが性能ネックになるという問題があった。第３の従来技術は、ディスクの回転待ち時間の影響を低減できる技術であり、フロントエンドとバックエンドのスループット乖離は低減できないという問題があった。

【０００８】本発明の目的は、ディスクアダプタとディスクアレイ間のスループットが高いディスク装置を提供することにある。本発明の他の目的は、ディスクアダプタとディスクアレイ間のスループットが高く、且つディスクドライブ接続台数が多いディスク装置を提供することにある。本発明のさらに他の目的は、信頼性が高いディスクアレイを有するディスク装置を提供することである。本発明のさらに他の目的は、信頼性が高いディスクアダプタとディスクアレイ間ネットワークを有するディスク装置を提供することにある。本発明のさらに他の目的は、信頼性およびスループットが高いディスクアダプタとディスクアレイ間ネットワークを有するディスク装置を提供することにある。本発明のさらに他の目的は、ディスクからの読み出しおよびディスクへの書きこみを高スループット化できるディスク装置を提供することにある。本発明のさらに他の目的は、高スループットを維持できるディスク装置を提供することである。本発明のさらに他の目的は、高スループットで低コストなディスク装置を提供することである。

【０００９】

【課題を解決するための手段】上記目的を達成するため、本発明は、ディスクコントローラとディスクアレイからなり、ディスクコントローラはチャネルアダプタとキャッシュメモリとディスクアダプタを有するディスク装置であり、ディスクアダプタとディスクアレイを、バッファメモリを有するスイッチを介して接続し、ディスクアダプタとスイッチ間のチャネル当りデータ転送速度を、スイッチとディスクアレイ間のチャネル当りデータ転送速度より高く設定し、スイッチは、ディスクアダプタが接続されたポートとディスクアレイを構成するディスクドライブが接続された各ポートとの間でのポート間の接続の切り換えを、入力されたフレーム毎に、該フレーム内の送信先情報にしたがって行っている。また、前記ディスクアレイはループ状に接続した複数のディスクドライブからなり、前記ディスクアダプタと前記複数のディスクアレイとをバッファメモリを有するスイッチを介して接続し、ディスクアダプタとスイッチ間のチャネル当りデータ転送速度を、スイッチと複数のディスクアレイ間のチャネル当りデータ転送速度より高く設定し、

スイッチは、ディスクアダプタが接続されたポートと複数のディスクアレイが接続された各ポートとの間でのポート間の接続の切り換えを、入力されたフレーム毎に、該フレーム内の送信先情報にしたがって行っている。また、前記ディスクアダプタと前記ディスクアレイを、バッファメモリを有するスイッチを介して接続し、同一のスイッチに接続したディスクドライブの組み合わせでＲＡＩＤグループを構成し、ディスクアダプタとスイッチ間のチャネル当りデータ転送速度を、スイッチとディスクアレイ間のチャネル当りデータ転送速度より高く設定し、スイッチは、ディスクアダプタが接続されたポートとＲＡＩＤグループを構成するディスクドライブが接続された各ポートとの間でのポート間の接続の切り換えを、入力されたフレーム毎に、該フレーム内の送信先情報にしたがって行っている。また、第１のディスクコントローラと第２のディスクコントローラと複数のディスクアレイからなり、第１のディスクコントローラは第１のチャネルアダプタと第１のキャッシュメモリと第１のディスクアダプタを有し、第２のディスクコントローラは第２のチャネルアダプタと第２のキャッシュメモリと第２のディスクアダプタを有するディスク装置であり、第１のディスクアダプタと前記複数のディスクアレイとをバッファメモリを有する第１のスイッチを介して接続し、且つ第２のディスクアダプタと前記複数のディスクアレイとをバッファメモリを有する第２のスイッチを介して接続し、さらに第１のスイッチと第２のディスクアダプタを接続し、第２のスイッチと第１のディスクアダプタを接続し、第２のディスクアダプタと第２のスイッチ間、および第１のディスクアダプタと第２のスイッチ間のチャネル当りデータ転送速度を第２のスイッチと前記複数のディスクアレイ間のチャネル当りデータ転送速度より高く設定し、第１のスイッチは、第１のディスクアダプタまたは第２のディスクアダプタが接続されたポートと前記複数のディスクアレイが接続された各ポートとの間でのポート間の接続の切り換えを、入力されたフレーム毎に、該フレーム内の送信先情報にしたがって行い、第２のスイッチは、第１のディスクアダプタまたは第２のディスクアダプタが接続されたポートと前記複数のディスクアレイが接続された各ポートとの間でのポート間の接続の切り換えを、入力されたフレーム毎に、該フレーム内の送信先情報にしたがって行っている。また、さらに、上記第１のスイッチと第２のスイッチを、上記第１のディスクアダプタと第２のスイッチ間を接続したチャネルと同等のデータ転送速度を有するチャネルと、第２のディスクアダプタと第１のスイッチ間を接続したチャネルと同等のデータ転送速度を有するチャネルと、を介して接続している。また、前記ディスクアレイからのデータ読み出し時には、前記ディスクアレイから前記スイッチに転送されるデータを前記スイッチにおいて多重化して前記ディスクアダプタに転送し、前記ディスクアレイへのデータ書き込み時には、前記ディスクアダプタから前記スイッチに転送されるデータを前記スイッチにおいて逆多重化して前記ディスクアレイに転送するようにしている。また、前記ディスクアダプタから前記ディスクアレイへのデータ書き込み時に、前記ディスクアダプタは、前記ポート間の接続の切り替えが行われるように、送出するフレームに送信先情報を設定し、前記ディスクアレイから前記ディスクアダプタへのデータ読み出し時に、前記スイッチは、ラウンドロビン方式により前記ポート間の接続を切り替えるようにしている。また、さらに、切り替えるポート数を、ディスクアダプタとスイッチ間のチャネル当りデータ転送速度の、スイッチとディスクアレイ間のチャネル当りデータ転送速度に対する比、と同程度に設定している。また、前記ディスクアダプタと前記スイッチ間を光ファイバケーブルで接続し、前記スイッチと前記ディスクアレイ間をメタルケーブルで接続するようにしている。

【００１０】

【発明の実施の形態】以下、図面を参照して本発明の実施の形態を詳細に説明する。図１に本発明の、第１の実施の形態であるディスク装置の構成を示す。本実施の形態のディスク装置は、ディスクコントローラＤＫＣとディスクアレイＤＡからなる。ディスクコントローラＤＫＣは、チャネルアダプタＣＨＡと、キャッシュメモリＣＭと、ディスクアダプタＤＫＡからなる。チャネルアダプタＣＨＡは、上位ＣＰＵ（図示せず）とディスクコントローラＤＫＣとがデータを送受信する際の制御を行う。Ｃ１、Ｃ２、Ｃ３およびＣ４は、チャネルアダプタＣＨＡがＣＰＵと通信するチャネルである。キャッシュメモリＣＭは、本実施の形態のディスク装置が入出力するデータを一時保存するメモリである。ディスクアダプタＤＫＡは、ディスクコントローラＤＫＣとディスクアレイＤＡとがデータを送受信する際の制御を行う。ディスクアダプタＤＫＡは、チャネルＤ０１、Ｄ０２、Ｄ０３、Ｄ０４を介して、ディスクアレイＤＡと接続する。ディスクアダプタＤＫＡとディスクアレイＤＡは、チャネルＤ０１、Ｄ０２、Ｄ０３、Ｄ０４上で全二重通信が可能である。

【００１１】ここで、本実施の形態のディスク装置は、ディスクアダプタＤＫＡとディスクアレイＤＡを、スイッチＳＷ１、ＳＷ２、ＳＷ３、ＳＷ４を介して接続している点に特徴がある。ディスクアレイＤＡは、ディスクグループＲ１、Ｒ２、Ｒ３、Ｒ４からなる。ディスクグループＲ１は、スイッチＳＷ１介してディスクアダプタＤＫＡと接続する。同様に、ディスクグループＲ２はスイッチＳＷ２を介して、ディスクグループＲ３はスイッチＳＷ３を介して、ディスクグループＲ４はスイッチＳＷ４介して、それぞれディスクアダプタＤＫＡと接続する。

【００１２】本実施の形態のディスク装置においてＲＡＩＤシステムを構築する場合は、ディスクグループＲ１、Ｒ２、Ｒ３、Ｒ４を、それぞれＲＡＩＤグループとする。本実施の形態では、４個のディスクドライブでＲＡＩＤグループを構成しているが、ＲＡＩＤグループを構成するドライブ数を４個に限るものではない。各ディスクグループへのデータ読み出しまたは書き込み時のデータの流れを、ディスクグループＲ１を例にして述べる。ここでＲ１はＲＡＩＤレベル５のＲＡＩＤグループである。チャネルＣ１、Ｃ２、Ｃ３、Ｃ４からディスクグループＲ１へ書き込むためにＣＰＵから送信されたデータは、ディスクアダプタＤＫＡにおいてブロック単位に分割されると同時に、該ブロック単位に分割されたデータからパリティが生成される。該ブロック単位に分割されたデータと、生成されたパリティは、チャネルＤ０１を通りスイッチＳＷ１へ入力される。スイッチＳＷ１は、ＲＡＩＤ制御に伴い、該ブロック単位に分割されたデータと、生成されたパリティとをルーティングし、チャネルＤ１１、Ｄ１２、Ｄ１３、Ｄ１４へ分配する。データ読み出し時は、ディスクアダプタＤＫＡは、Ｄ１１、Ｄ１２、Ｄ１３、Ｄ１４を介してディスクグループＲ１からブロック単位に分割されたデータを読み出し、スイッチＳＷ１でシリアル化して、チャネルＤ０１を介して読み出しデータを受信する。

【００１３】図９に示した従来のディスク装置では、ディスクアダプタＤＫＡに接続したチャネルＤ１、Ｄ２、Ｄ３、Ｄ４上で、既にディスクアレイへの書き込みデータおよびパリティが別々のチャネルに分配されていた。それに対し、本実施の形態のディスク装置においては、スイッチＳＷ１通過後に別々のチャネルに分配される点が従来と異なる。

【００１４】次に、本実施の形態のディスク装置の特徴であるスイッチの動作を、スイッチＳＷ１を例にとり説明する。ＳＷ２～ＳＷ４の動作もＳＷ１の動作と同様である。図１に示すように、スイッチＳＷ１は入出力ポートＰ１、Ｐ２、Ｐ３、Ｐ４、Ｐ５を有する。ポートＰ１、Ｐ２、Ｐ３、Ｐ４、Ｐ５は、全二重通信可能な入出力ポートであり、ポート毎にバッファメモリを有している。スイッチＳＷ１の内部構成を図２と図３に示す。簡単のため、データの進行方向によりスイッチ動作を分けて説明する。また、チャネルＤ０１、Ｄ１１、Ｄ１２、Ｄ１３、Ｄ１４上を流れるデータは、フレーム単位で送受信され、かつデータは８Ｂ１０Ｂ変換で符号化されている。

【００１５】図２は、ポートＰ１からブロック内のフレームを入力し、ポートＰ２、Ｐ３、Ｐ４、Ｐ５から出力する場合を示す。これはディスクアレイへの書き込み時のスイッチ動作に相当する。スイッチＳＷ１は図２に示すように、クロスバスイッチＸＳＷと、スイッチコントローラＣＴＬからなる。クロスバスイッチＸＳＷは５×５のクロスバスイッチであり、入力ポートｉｎ１、ｉｎ２、ｉｎ３、ｉｎ４、ｉｎ５と、出力ポートｏｕｔ１、ｏｕｔ２、ｏｕｔ３、ｏｕｔ４、ｏｕｔ５を有する。ポートＰ１から入力したフレームは、シリアルパラレル変換装置ＳＰ１と、バッファメモリＢＭ１と、８Ｂ１０Ｂ変換デコーダＤＥＣ１を経由し、スイッチコントローラＣＴＬと入力ポートｉｎ１へ入力される。スイッチコントローラＣＴＬにおいて、入力フレームのヘッダ部分に書かれた送信先アドレスを解読し、クロスバスイッチＸＳＷを切り換える。例として、ポートＰ２が出力先として選ばれた場合は、入力したフレームは出力ポートｏｕｔ２と、８Ｂ１０Ｂ変換エンコーダＥＮＣ２と、バッファメモリＢＭ２と、パラレルシリアル変換装置ＰＳ２を経由して、ポートＰ２から出力される。ここで、バッファメモリＢＭ１、ＢＭ２はＦＩＦＯ（Ｆｉｒｓｔ－ｉｎ　Ｆｉｒｓｔ－ｏｕｔ）メモリである。

【００１６】シリアルパラレル変換装置ＳＰ１は、８Ｂ１０Ｂ符号化されたシリアルデータを１０ｂｉｔ幅のパラレルデータに変換し、ポートＰ１におけるデータ転送速度の１／１０の速度に同期してバッファメモリＢＭ１に書き込む。８Ｂ１０Ｂデコーダ ＤＥＣ１は、クロスバスイッチＸＳＷの動作速度に同期して、１０ｂｉｔパラレルデータをバッファメモリＢＭ１から読み出し、８Ｂ１０Ｂ復号化して、８ｂｉｔパラレルデータに変換する。８Ｂ１０Ｂエンコーダ ＥＮＣ２は、クロスバスイッチＸＳＷでスイッチされた８ｂｉｔパラレルデータを再び８Ｂ１０Ｂ符号化し、１０ｂｉｔパラレルデータに変換後、クロスバスイッチＸＳＷの動作速度に同期してバッファメモリＢＭ２に書き込む。パラレルシリアル変換装置ＰＳ２は、ポートＰ２におけるデータ転送速度の１／１０の速度に同期して、１０ｂｉｔパラレルデータをバッファメモリＢＭ２から読み出し、シリアル化して、ポートＰ２から出力する。以上によりスイッチＳＷ１は、ポートＰ１におけるデータ転送速度からポートＰ２におけるデータ転送速度へ速度変換する。

【００１７】図４は、ポートＰ１へ入力するフレームと、ポートＰ２、Ｐ３、Ｐ４、Ｐ５から出力されるフレームを示した図である。波形の凸はフレームが存在する時間、凹はフレームが存在していない時間を示している。フレームは伝送するデータ容量に従ってそのフレーム長が変化するが、ここではディスクアレイへのシーケンシャルアクセスが行われており、フレーム長が一定である。図４では、入力ポートＰ１でのデータ転送速度が出力ポートＰ２、Ｐ３、Ｐ４、Ｐ５におけるデータ転送速度のｍ倍あるとする。従って、ポートＰ１におけるフレームＦｂ２の時間Ｔ１は、ポートＰ２からの出力時にＴ３へ伸びている。ここでＴ３＝ｍ×Ｔ１である。

【００１８】入力のデータ転送速度が速く、且つ出力のデータ転送速度が遅い場合は、スイッチを周期的に切り換えないと出力ポートのバッファメモリが溢れ、スループットが低下する。フレームがスループットの低下無く

スイッチを通過するには、図４のように周期的に出力ポートを切り換える必要がある。スイッチ切替えポート数をｎとすると、スイッチ切替え周期Ｔ２≒ｎ×Ｔ１である（フレームの無い時間は無視した）。Ｔ２≧Ｔ３ならば、フレームの衝突無く、スループットの低下は起こらない。Ｔ２≧Ｔ３はｎ≧ｍと同じある。つまり、ディスクアレイへのデータ書き込み時に、スイッチにおいてスループット低下を起こさないための条件は、周期的に切り替えるスイッチポート数ｎを、ディスクアダプタとスイッチ間のチャネル当りデータ転送速度の、スイッチとディスクアレイ間のチャネル当りデータ転送速度に対する比ｍ、以上に設定することである。この条件が保たれれば、スイッチＳＷ１は、ポートＰ１から入力したデータをバッファメモリにおいて速度変換し、フレーム単位で周期的に切り替えることにより逆多重化し、ポートＰ２、Ｐ３、Ｐ４、Ｐ５へ分配して出力する。スイッチを周期的に切り換える方法の一つは、スイッチに接続したディスクグループをＲＡＩＤグループとすることである。ＲＡＩＤのストライピング制御に従えば、スイッチは周期的に切り替わる。

【００１９】図３は、ポートＰ２、Ｐ３、Ｐ４、Ｐ５からフレームを入力し、ポートＰ１から出力する場合を示す。これはディスクアレイからの読み出し時のスイッチ動作に相当する。例えば、ポートＰ２から入力したフレームは、シリアルパラレル変換装置ＳＰ２と、バッファメモリＢＭ２と、８Ｂ１０Ｂ変換デコーダＤＥＣ２を経由し、スイッチコントローラＣＴＬと入力ポートｉｎ２へ入力される。スイッチコントローラＣＴＬにおいて、入力フレームのヘッダ部分に書かれた送信先アドレスを解読し、クロスバスイッチＸＳＷを切り換える。図３の場合は、ラウンドロビン方式によりクロスバススイッチＸＳＷを切り替えて、順番にポートＰ２、Ｐ３、Ｐ４、Ｐ５から入力されるデータは全てポートＰ１へ出力する。すなわち、読み出し時は、複数の入力ポート（Ｐ２、Ｐ３、Ｐ４、Ｐ５）に同時にフレームが届く。これら複数の入力フレームは同期して入力ポートに届く必要はない。スイッチは、総当り的に入出力ポート間接続を切り替えることにより、これら複数の入力フレームを１フレームずつ出力ポート（Ｐ１）へ転送する。このように、スイッチを総当り的に切り替える方式を、ラウンドロビン(Round Robin)方式と呼ぶ。ラウンドロビン方式により、結果的にスイッチは周期的に切り替わることになる。なお、読み出し時においても、スイッチはフレーム内送信先情報に従って切り替わることに違いはない。フレームは出力ポートｏｕｔ１と、８Ｂ１０Ｂ変換エンコーダＥＮＣ１と、バッファメモリＢＭ１と、パラレルシリアル変換装置ＰＳ１を経由して、ポートＰ１から出力される。

【００２０】シリアルパラレル変換装置ＳＰ２は、８Ｂ１０Ｂ符号化されたシリアルデータを１０ｂｉｔ幅のパラレルデータに変換し、ポートＰ２におけるデータ転送速度の１／１０の速度に同期してバッファメモリＢＭ２に書き込む。８Ｂ１０Ｂデコーダ ＤＥＣ２は、クロスバスイッチＸＳＷの動作速度に同期して、１０ｂｉｔパラレルデータをバッファメモリＢＭ２から読み出し、８Ｂ１０Ｂ復号化して、８ｂｉｔパラレルデータに変換する。８Ｂ１０ＢエンコーダＥＮＣ１は、クロスバスイッチＸＳＷでスイッチされた８ｂｉｔパラレルデータを再び８Ｂ１０Ｂ符号化し、１０ｂｉｔパラレルデータに変換後、クロスバスイッチＸＳＷの動作速度に同期してバッファメモリＢＭ１に書き込む。パラレルシリアル変換装置ＰＳ１は、ポートＰ１におけるデータ転送速度の１／１０の速度に同期して、１０ｂｉｔパラレルデータをバッファメモリＢＭ１から読み出し、シリアル化して、ポートＰ１から出力する。以上によりスイッチＳＷ１は、ポートＰ２におけるデータ転送速度からポートＰ１におけるデータ転送速度へ速度変換する。

【００２１】図５は、ポートＰ２、Ｐ３、Ｐ４、Ｐ５へ入力するフレームと、ポートＰ１から出力されるフレームを示した図である。波形の凸はフレームが存在する時間、凹はフレームが存在していない時間を示している。フレームは伝送するデータ容量に従ってそのフレーム長が変化するが、ここではディスクアレイへのシーケンシャルアクセスが行われており、フレーム長が一定である。図５では、入力ポートＰ１でのデータ転送速度が出力ポートＰ２、Ｐ３、Ｐ４、Ｐ５におけるデータ転送速度のｍ倍あるとする。従って、ポートＰ５におけるフレームＦｅ５の時間Ｔ４は、ポートＰ１からの出力時にＴ５へ縮んでいる。ここでＴ４＝ｍ×Ｔ５である。フレームＦｅ２、Ｆｅ３、Ｆｅ４、Ｆｅ５をポートＰ１から出力するのにかかる時間をＴ６とする。スイッチ切り替えポート数をｎとすると、Ｔ６≒ｎ×Ｔ５である（フレームの無い時間は無視した）。スイッチにおいて輻輳によるスループット低下を防止するためには、Ｔ６≦Ｔ４とする必要がある。Ｔ６≦Ｔ４はｎ≦ｍと同じある。

【００２２】つまり、ディスクアレイからのデータ読み出し時に、スイッチにおいてスループット低下を起こさないための条件は、周期的に切り替えるスイッチポート数ｎを、ディスクアダプタとスイッチ間のチャネル当りデータ転送速度の、スイッチとディスクアレイ間のチャネル当りデータ転送速度に対する比ｍ、以下に設定することである。この条件が保たれれば、スイッチＳＷ１は、ポートＰ２、Ｐ３、Ｐ４、Ｐ５から入力したデータをバッファメモリにおいて速度変換し、フレーム単位で周期的に切り替えることにより多重化し、ポートＰ１へ出力する。よって、ディスクアレイへの書き込みおよびディスクアレイからの読み出しを高スループット化するためには、ｎ≒ｍ、つまり、周期的に切り　えるポート数を、ディスクアダプタとスイッチ間のチャネル当りデータ転送速度の、スイッチとディスクアレイ間のチャネル当りデータ転送速度に対する比、と同程度に設定すればよいことが分かる。

【００２３】例えば、ディスクアダプタとスイッチ間の４Ｇｂｐｓのチャネル１本で接続し、スイッチとディスクアレイ間を１Ｇｂｐｓのチャネル４本で接続する。また、ディスクアダプタとスイッチ間の１０Ｇｂｐｓのチャネル１本で接続し、スイッチとディスクアレイ間を２Ｇｂｐｓのチャネル４本で接続する。この場合、スイッチ入出力ポート間でスループットのバランスが取れないので、実効的なスループットは２Ｇｂｐｓ×４＝８Ｇｂｐｓとなる。

【００２４】以上より、スイッチＳＷ１において速度変換と多重化、逆多重化が行われるので、チャネルＤ１１、Ｄ１２、Ｄ１３、Ｄ１４上のデータ転送速度が低速でも、チャネルＤ０１、Ｄ０２、Ｄ０３、Ｄ０４でのデータ転送速度は高速にできる。つまり、ディスクアダプタＤＫＡとディスクアレイＤＡ間のスループットを向上できる。本実施の形態のディスク装置におけるデータ転送方式としては、ファイバチャネルやインフィニバンドが使用できる。

【００２５】図６は、第１の実施の形態のディスク装置において、ディスクドライブの増設方法を示した図である。図６では図１に対して、ディスクグループＲ５とＲ６が増設されている。ディスクドライブを増設するため、スイッチＳＷ１とＳＷ２としてポート数の多いスイッチを使用している。ディスクドライブを増設すると、スイッチのディスクアレイ側のスループットが増加し、ディスクアダプタ側とのスループットバランスが崩れるので、スイッチの速度変換機能が有効に働かなくなる可能性がある。そこでスイッチＳＷ１では、ディスクアダプタＤＫＡとの間に、新規チャネルＤ０５を増設している。また、スイッチＳＷ２の場合は新規チャネルを増設せず、チャネルＤ０２の信号伝送速度を増加させることで、ディスクアダプタ側とディスクアレイ側のスループットバランスを取っている。例えばスイッチＳＷ１では、スイッチとディスクアレイ間を１Ｇｂｐｓのチャネル８本で接続し、ディスクアダプタとスイッチ間を４Ｇｂｐｓのチャネル２本で接続する。スイッチＳＷ２では、スイッチとディスクアレイ間を１Ｇｂｐｓのチャネル８本で接続し、ディスクアダプタとスイッチ間を１０Ｇｂｐｓのチャネル１本で接続する。このように、本実施の形態のディスク装置は、スイッチのポート数に応じて、ディスクドライブを増設可能である。このディスクドライブ増設方法は、１ポート当たりに接続できるドライブ数が少ないＡＴＡ（ＡＴ　Ａｔｔａｃｈｍｅｎｔ）方式ディスクドライブを増設するのに適用できる。

【００２６】図７に本発明の、第２の実施の形態であるディスク装置の構成を示す。本実施の形態のディスク装置は、第１の実施の形態のディスク装置に対して、ディスクアレイ部分の構成方法が異なる。本実施の形態のディスク装置は、ディスクコントローラＤＫＣと、４個のディスクアレイＤＡ１、ＤＡ２、ＤＡ３、ＤＡ４からなる。ディスクコントローラＤＫＣは、チャネルアダプタＣＨＡ、キャッシュメモリＣＭ、ディスクアダプタＤＫＡからなる。ディスクアレイＤＡ１とディスクアダプタＤＫＡは、チャネルＤ０１とスイッチＳＷ１を介して接続する。同様に、ディスクアレイＤＡ２はチャネルＤ０２とスイッチＳＷ２を介して、ディスクアレイＤＡ３はチャネルＤ０３とスイッチＳＷ３を介して、ディスクアレイＤＡ４はチャネルＤ０４とスイッチＳＷ４を介して、それぞれディスクアダプタＤＫＡと接続する。スイッチＳＷ１、ＳＷ２、ＳＷ３とＳＷ４は、第１の実施の形態と同様に速度変換と多重化、逆多重化を行うスイッチとして機能する。本実施の形態におけるディスクアダプタＤＫＡと、スイッチＳＷ１、ＳＷ２、ＳＷ３、ＳＷ４と、ディスクアレイＤＡ１、ＤＡ２、ＤＡ３、ＤＡ４との間のデータ転送方式は、ファイバチャネルを使用している。スイッチＳＷ１、ＳＷ２、ＳＷ３、ＳＷ４はファイバチャネルスイッチである。

【００２７】本実施の形態におけるディスクアレイの構成を、ディスクアレイＤＡ１を例に述べる。ディスクアレイＤＡ１、ＤＡ２、ＤＡ３、ＤＡ４は、同様のドライブ構成である。ディスクアレイＤＡ１は、チャネルＤ１１上に接続した４個のディスクからなるディスクアレイと、Ｄ１２上に接続した４個のディスクからなるディスクアレイと、Ｄ１３上に接続した４個のディスクからなるディスクアレイと、Ｄ１４上に接続した４個のディスクからなるディスクアレイ、からなる。チャネルＤ１１を例にとると、ディスクドライブＤＫ１、ＤＫ２、ＤＫ３、ＤＫ４が、チャネルＤ１１上に接続されている。このように、多数のドライブを一つのチャネル上に接続してディスクドライブにアクセスする方法としては、ファイバチャネルアービトレイテッドループ（以下ＦＣ－ＡＬと呼ぶ）がある。

【００２８】図１０に、ＦＣ－ＡＬの接続形態をディスクドライブＤＫ１、ＤＫ２、ＤＫ３、ＤＫ４の接続形態を例として示す。各ディスクドライブの入出力ポートおよびスイッチＳＷ１の入出力ポートは、送信機Ｔｘと受信機Ｒｘを有する。ＦＣ－ＡＬの接続形態は、例えば図１０に示すように、各ドライブの入出力ポートおよびスイッチの入出力ポートをループ状に接続するトポロジである。各ドライブの入出力ポートはファイバチャネルのＮＬ（Ｎｏｄｅ　Ｌｏｏｐ）ポートとして機能する。ＮＬポートとは、ループ動作をする装置（ここではディスクドライブ）のポートである。スイッチＳＷ１のディスクアレイＤＡ１接続側入出力ポートは、ファイバチャネルのＦＬ（Ｆａｂｒｉｃ　Ｌｏｏｐ）ポートとして機能する。ＦＬポートとは、ＦＣ－ＡＬを接続可能なスイッチのポートである。ＦＬポートを有するループは、ファイバチャネルのパブリックループとして機能するので、

チャネルＤ１１が形成するＦＣ－ＡＬはパブリックループとなる。パブリックループとは、ループ上のディスクドライブが、スイッチを介してループ外のポートと通信可能なループである。よって、ディスクドライブＤＫ１、ＤＫ２、ＤＫ３、ＤＫ４は、スイッチＳＷ１およびチャネルＤ０１を介してディスクアダプタＤＫＡと通信可能である。以上、チャネルＤ１１の接続形態を例に説明したが、チャネルＤ１２、Ｄ１３、Ｄ１４でも同様である。本実施の形態のディスク装置においてＲＡＩＤシステムを構築する場合は、ディスクグループＲ１、Ｒ２、Ｒ３、Ｒ４を、それぞれＲＡＩＤグループとする。本実施の形態では、４個のディスクドライブでＲＡＩＤグループを構成しているが、ＲＡＩＤグループを構成するドライブ数を４個に限るものではない。

【００２９】本実施の形態においては、チャネルＤ１１、Ｄ１２、Ｄ１３、Ｄ１４において、それぞれＦＣ－ＡＬを用いてディスクドライブを接続している。ＦＣ－ＡＬの仕様から、チャネルＤ１１、Ｄ１２、Ｄ１３、Ｄ１４上には、それぞれ最大１２６台までのディスクドライブが接続可能である。また、チャネルＤ０１、Ｄ０２、Ｄ０３、Ｄ０４の媒体として光ファイバケーブルを、チャネルＤ１１、Ｄ１２、Ｄ１３、Ｄ１４の媒体としてメタルケーブルを用いる。

【００３０】以上説明したように、本実施の形態のディスク装置においては、ディスクドライブをＦＣ－ＡＬで接続しているので、スイッチのポート当りに接続できるドライブ台数が増加できる。つまり、ディスク装置当りの記憶容量を増加させる効果がある。また、ディスクドライブをメタルケーブルで接続することにより、ディスクドライブ毎に高価な光インターフェイスを装備する必要がなくなるので、ディスクドライブのコストを下げる効果がある。

【００３１】図８に本発明の、第３の実施の形態であるディスク装置の構成を示す。本実施の形態のディスク装置は、ディスクコントローラとスイッチを二重化した点に特徴がある。本実施の形態において、ディスクアダプタＤＫＡ１、ＤＫＡ２と、スイッチＳＷ１、ＳＷ２と、ディスクアレイＤＡ１との間のデータ転送方式は、ファイバチャネルを使用している。本実施の形態のディスク装置は、ディスクコントローラＤＫＣ１、ＤＫＣ２と、スイッチＳＷ１、ＳＷ２と、ディスクアレイＤＡ１からなる。スイッチＳＷ１とＳＷ２は、第１の実施の形態と同様に速度変換と多重化、逆多重化を行うスイッチとして機能する。ディスクコントローラＤＫＣ１は、チャネルアダプタＣＨＡ１と、キャッシュメモリＣＭ１と、ディスクアダプタＤＫＡ１からなる。ディスクコントローラＤＫＣ２は、チャネルアダプタＣＨＡ２と、キャッシュメモリＣＭ２と、ディスクアダプタＤＫＡ２からなる。ディスクアダプタＤＫＡ１とスイッチＳＷ１をチャネルＤ１ａで接続し、ディスクアダプタＤＫＡ２とスイッチＳＷ２をチャネルＤ２ａで接続し、ディスクアダプタＤＫＡ１とスイッチＳＷ２をチャネルＤ１ｂで接続し、ディスクアダプタＤＫＡ２とスイッチＳＷ１をチャネルＤ２ｂで接続する。

【００３２】ディスクアレイＤＡ１を構成するディスクドライブは、入出力ポートを２個有する。例えば、ディスクドライブＤＫ１、ＤＫ２、ＤＫ３、ＤＫ４は、チャネルＤ１１およびＤ２１の両チャネルと接続する。ディスクアレイＤＡ１は、チャネルＤ１１とＤ２１に接続した４個のディスクからなるディスクアレイと、Ｄ１２とＤ２２に接続した４個のディスクからなるディスクアレイと、Ｄ１３とＤ２３に接続した４個のディスクからなるディスクアレイと、Ｄ１４とＤ２４に接続した４個のディスクからなるディスクアレイ、からなる。チャネルＤ１１、Ｄ１２、Ｄ１３、Ｄ１４、Ｄ２１、Ｄ２２、Ｄ２３、Ｄ２４は、ＦＣ－ＡＬでディスクドライブを接続する。

【００３３】図１１に本実施の形態におけるＦＣ－ＡＬの接続形態を、ディスクドライブＤＫ１、ＤＫ２、ＤＫ３、ＤＫ４の接続形態を例として示す。各ディスクドライブは、それぞれＮＬポートを２個有する。各ディスクドライブの入出力ポートおよびスイッチＳＷ１、ＳＷ２の入出力ポートは、送信機Ｔｘと受信機Ｒｘを有する。スイッチＳＷ１、ＳＷ２のディスクアレイＤＡ１接続側入出力ポートは、ＦＬポートである。チャネルＤ１１により、スイッチＳＷ１、ディスクドライブＤＫ１、ＤＫ２、ＤＫ３、ＤＫ４をループ状に接続する。同様にチャネルＤ２１により、スイッチＳＷ２、ディスクドライブＤＫ１、ＤＫ２、ＤＫ３、ＤＫ４をループ状に接続する。これら２個のループは、ファイバチャネルのパブリックループであり、ディスクドライブＤＫ１、ＤＫ２、ＤＫ３、ＤＫ４は、スイッチＳＷ１またはＳＷ２を介してディスクアダプタＤＫＡ１またはＤＫＡ２と通信可能である。以上、チャネルＤ１１、Ｄ２１の接続形態を例に説明したが、チャネルＤ１２、Ｄ１３、Ｄ１４、Ｄ２２、Ｄ２３、Ｄ２４でも同様である。本実施の形態のディスク装置においてＲＡＩＤシステムを構築する場合は、ディスクグループＲ１、Ｒ２、Ｒ３、Ｒ４を、それぞれＲＡＩＤグループとする。本実施の形態では、４個のディスクドライブでＲＡＩＤグループを構成しているが、ＲＡＩＤグループを構成するドライブ数を４個に限るものではない。

【００３４】ディスクアレイＤＡ１内の全ディスクドライブは、ディスクアダプタＤＫＡ１およびＤＫＡ２のどちらからでもアクセス可能である。本実施の形態のディスク装置は、チャネルＤ１ｂ、Ｄ２ｂをスイッチＳＷ１、ＳＷ２故障時の迂回経路として使用する。例えばスイッチＳＷ１が故障した場合でも、ディスクアダプタＤＫＡ１はチャネルＤ１ｂとスイッチＳＷ２経由でディスクアレイＤＡ１にアクセスできる。逆に、スイッチＳＷ２が故障した場合は、ディスクアダプタＤＫＡ２はチャネルＤ２ｂとスイッチＳＷ１経由でディスクアレイＤＡ１にアクセスできるので、信頼性が高いディスク装置が実現できる。

【００３５】図１２に本発明の、第４の実施の形態であるディスク装置の構成を示す。本実施の形態のディスク装置は、第３の実施の形態のディスク装置に対して、スイッチＳＷ１、ＳＷ２間を接続するチャネルＤ３ａ、Ｄ３ｂを設けた点に特徴が有る。本実施の形態において、ディスクアダプタＤＫＡ１、ＤＫＡ２と、スイッチＳＷ１、ＳＷ２と、ディスクアレイＤＡ１との間のデータ転送方式は、ファイバチャネルを使用している。本実施の形態のディスク装置は、ディスクコントローラＤＫＣ１、ＤＫＣ２と、スイッチＳＷ１、ＳＷ２と、ディスクアレイＤＡ１からなる。スイッチＳＷ１とＳＷ２は、第１の実施の形態と同様に速度変換と多重化、逆多重化を行うスイッチとして機能する。ディスクコントローラＤＫＣ１は、チャネルアダプタＣＨＡ１と、キャッシュメモリＣＭ１と、ディスクアダプタＤＫＡ１からなる。スイッチＳＷ１とＳＷ２をチャネルＤ１ａで接続し、ディスクアダプタＤＫＡ２とスイッチＳＷ２をチャネルＤ２ａで接続し、ディスクアダプタＤＫＡ１とスイッチＳＷ２をチャネルＤ１ｂで接続し、ディスクアダプタＤＫＡ２とスイッチＳＷ１をチャネルＤ２ｂで接続する。スイッチＳＷ１とＳＷ２をチャネルＤ３ａ、Ｄ３ｂで接続する。

【００３６】ディスクアレイＤＡ１を構成するディスクドライブは、入出力ポートを２個有する。例えば、ディスクドライブＤＫ１、ＤＫ２、ＤＫ３、ＤＫ４は、チャネルＤ１１およびＤ２１の両チャネルと接続する。ディスクアレイＤＡ１は、チャネルＤ１１とＤ２１に接続した４個のディスクからなるディスクアレイと、Ｄ１２とＤ２２に接続した４個のディスクからなるディスクアレイと、Ｄ１３とＤ２３に接続した４個のディスクからなるディスクアレイと、Ｄ１４とＤ２４に接続した４個のディスクからなるディスクアレイ、からなる。チャネルＤ１１、Ｄ１２、Ｄ１３、Ｄ１４、Ｄ２１、Ｄ２２、Ｄ２３、Ｄ２４は、図１１に示すようにＦＣ－ＡＬでディスクドライブを接続する。ディスクアレイＤＡ１内の全ディスクドライブは、ディスクアダプタＤＫＡ１およびＤＫＡ２のどちらからでもアクセス可能である。本実施の形態のディスク装置においてＲＡＩＤシステムを構築する場合は、ディスクグループＲ１、Ｒ２、Ｒ３、Ｒ４を、それぞれＲＡＩＤグループとする。本実施の形態では、４個のディスクドライブでＲＡＩＤグループを構成しているが、ＲＡＩＤグループを構成するドライブ数を４個に限るものではない。

【００３７】ディスクアダプタＤＫＡ１、ＤＫＡ２とディスクアレイＤＡ１のアクセス経路について、先ず、定常時（スイッチ故障無しの場合）について説明する。ディスクアダプタＤＫＡ１は、チャネルＤ１ａとＳＷ１を介してディスクアレイＤＡ１にアクセスする経路（経路１）と、チャネルＤ１ｂとスイッチＳＷ２とチャネルＤ３ａとスイッチＳＷ１を介してディスクアレイＤＡ１にアクセスする経路（経路２）を有する。同様に、ディスクアダプタＤＫＡ２は、チャネルＤ２ａとＳＷ２を介してディスクアレイＤＡ１にアクセスする経路（経路３）と、チャネルＤ２ｂとスイッチＳＷ１とチャネルＤ３ｂとスイッチＳＷ２を介してディスクアレイＤＡ１にアクセスする経路（経路４）を有する。一方、スイッチ故障時は、チャネルＤ１ｂ、Ｄ２ｂを迂回経路として使用する。例えばスイッチＳＷ１が故障した場合でも、ディスクアダプタＤＫＡ１はチャネルＤ１ｂとスイッチＳＷ２経由でディスクアレイＤＡ１にアクセスできる。逆に、スイッチＳＷ２が故障した場合は、ディスクアダプタＤＫＡ２はチャネルＤ２ｂとスイッチＳＷ１経由でディスクアレイＤＡ１にアクセスできる。

【００３８】次に、本実施の形態におけるディスクアダプターディスクアレイ間のスループットについて説明する。例として、チャネルＤ１ａ、Ｄ１ｂ、Ｄ２ａ、Ｄ２ｂ、Ｄ３ａ、Ｄ３ｂ上のデータ伝送速度をチャネル当り２Ｇｂｐｓとし、チャネルＤ１１、Ｄ１２、Ｄ１３、Ｄ１４、Ｄ２１、Ｄ２２、Ｄ２３、Ｄ２４上のデータ伝送速度をチャネル当り１Ｇｂｐｓであるとする。このとき、スイッチＳＷ１とディスクアレイＤＡ１間の総スループットは４Ｇｂｐｓである。ディスクアダプタＤＫＡ１とスイッチＳＷ１間は、上記経路１および経路２でアクセスすることにより、総スループットは４Ｇｂｐｓとなる。スイッチＳＷ１のディスクアダプタＤＫＡ１側と、ディスクアレイＤＡ１側のスループットが共に４Ｇｂｐｓであるので、ディスクアダプタＤＫＡ１とディスクアレイＤＡ１間のスループットは４Ｇｂｐｓとなる。同様に、スイッチＳＷ２とディスクアレイＤＡ１間の総スループットは４Ｇｂｐｓである。ディスクアダプタＤＫＡ２とスイッチＳＷ２間は、上記経路３および経路４でアクセスすることにより、総スループットは４Ｇｂｐｓとなる。スイッチＳＷ２のディスクアダプタＤＫＡ２側と、ディスクアレイＤＡ１側のスループットが共に４Ｇｂｐｓであるので、ディスクアダプタＤＫＡ２とディスクアレイＤＡ２間のスループットは４Ｇｂｐｓとなる第３の実施の形態（図８）において、上記のチャネル当りスループット値を適用すると、チャネルＤ１ｂ、Ｄ２ｂをスイッチ故障時の迂回経路としか使用しないので、ディスクアダプタＤＫＡ１とディスクアレイＤＡ１間のスループットは、チャネルＤ１ａ上のスループットに制限され、２Ｇｂｐｓとなる。同様に、ディスクアダプタＤＫＡ２とディスクアレイＤＡ１間のスループットは、チャネルＤ２ａ上のスループットに制限され、２Ｇｂｐ

ｓとなる。第３の実施の形態において、ディスクアダプターディスクアレイ間スループットを４Ｇｂｐｓにするためには、チャネルＤ１ａおよびＤ２ａのデータ伝送速度を、それぞれ４Ｇｂｐｓに高める必要がある。以上から、本実施の形態によれば、ディスクアダプタースイッチ間のチャネル当りデータ伝送速度が低くても、ディスクアダプターディスクアレイ間の総スループットが高いディスク装置が実現できる。

【００３９】

【発明の効果】以上説明したように、本発明によれば以下の効果がある。ディスクアダプタとディスクアレイ間のスループットが高いディスク装置を提供できる。また、ディスクアダプタとディスクアレイ間のスループットが高く、且つディスクドライブ接続台数が多いディスク装置を提供できる。また、信頼性の高いディスクアレイを有するディスク装置を提供できる。また、信頼性が高いディスクアダプタとディスクアレイ間ネットワークを有するディスク装置を提供できる。また、信頼性およびスループットが高いディスクアダプタとディスクアレイ間ネットワーク、を有するディスク装置を提供できる。また、ディスクからの読み出しおよびディスクへの書きこみを高スループット化できるディスク装置を提供できる。また、高スループットを維持できるディスク装置を提供できる。また、ディスクアダプタとディスクアレイ間のスループットが高く低コストなディスク装置を提供できる。

【図面の簡単な説明】

【図１】本発明の第１の実施の形態のディスク装置を示す図である。

【図２】本発明に用いるスイッチの構成を示す図である。

【図３】本発明に用いるスイッチの構成を示す図である。

【図４】本発明に用いるスイッチの動作を示す図である。

【図５】本発明に用いるスイッチの動作を示す図である。

【図６】本発明第１の実施の形態に対して、ディスクドライブを増設する方法を示す図である。

【図７】本発明の第２の実施の形態のディスク装置を示す図である。

【図８】本発明の第３の実施の形態のディスク装置を示す図である。

【図９】従来のディスク装置を示す図である。

【図１０】ＦＣ－ＡＬによる接続形態を説明する図である。

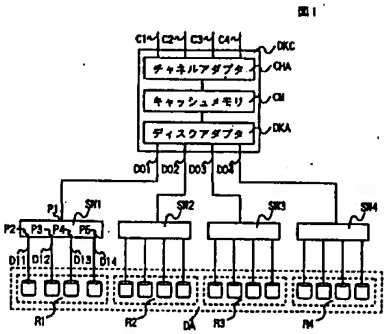
【図１１】ＦＣ－ＡＬによる接続形態を説明する図である。

【図１２】本発明の第４の実施の形態のディスク装置を示す図である。

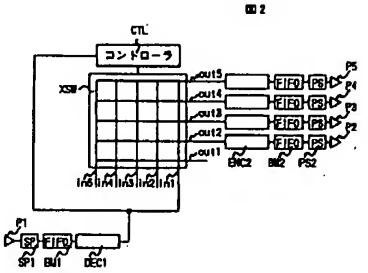
【符号の説明】

ＤＫＣ、ＤＫＣ１、ＤＫＣ２　ディスクコントローラ
ＣＨＡ、ＣＨＡ１、ＣＨＡ２　チャネルアダプタ
ＣＭ、ＣＭ１、ＣＭ２　キャッシュメモリ
ＤＫＡ、ＤＫＡ１、ＤＫＡ２　ディスクアダプタ
ＤＡ、ＤＡ１～ＤＡ４　ディスクアレイ
ＤＫ１～ＤＫ４　ディスクドライブ
Ｒ１～Ｒ６　ディスクグループ
Ｃ１～Ｃ４、Ｄ１～Ｄ４、Ｄ０１～Ｄ０５、Ｄ１１～Ｄ１４、Ｄ２１～Ｄ２４、Ｄ１ａ、Ｄ１ｂ、Ｄ２ａ、Ｄ２ｂ、Ｄ３ａ、Ｄ３ｂ　チャネル
ＳＷ１～ＳＷ４　スイッチ
Ｐ１～Ｐ５　スイッチポート
ＸＳＷ　クロスバスイッチ
ＣＴＬ　スイッチコントローラ
ｉｎ１～ｉｎ５　クロスバスイッチ入力ポート
ｏｕｔ１～ｏｕｔ５　クロスバスイッチ出力ポート
ＳＰ１、ＳＰ２　シリアルパラレル変換装置
ＰＳ１、ＰＳ２　パラレルシリアル変換装置
ＢＭ１、ＢＭ２　バッファメモリ
ＤＥＣ１、ＤＥＣ２　８Ｂ１０Ｂ変換デコーダ
ＥＮＣ１、ＥＮＣ２　８Ｂ１０Ｂ変換エンコーダ
Ｔ１、Ｔ２、Ｔ３、
Ｔ４、Ｔ５、Ｔ６　フレームの時間
Ｔｘ　送信機
Ｒｘ　受信機
ＮＬ　ＮＬポート
ＦＬ　ＦＬポート

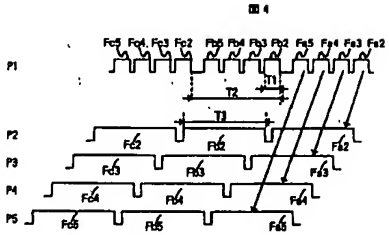
【図１】



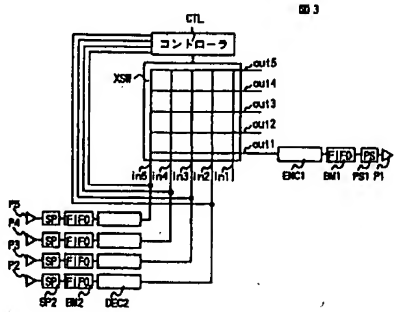
【図２】



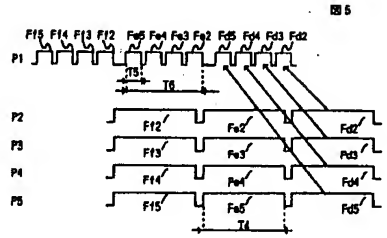
【図４】



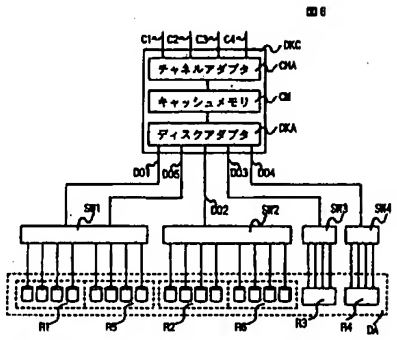
【図３】



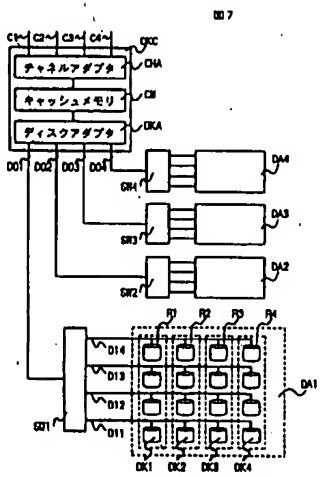
【図５】



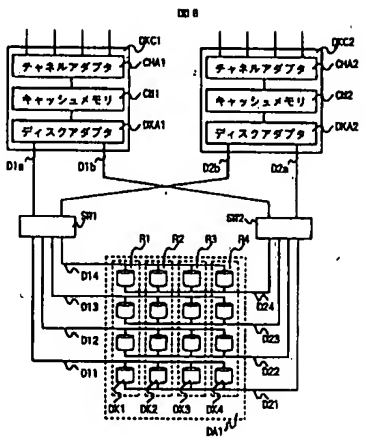
【図６】



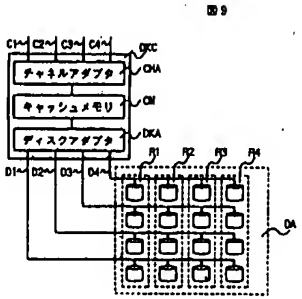
【図７】



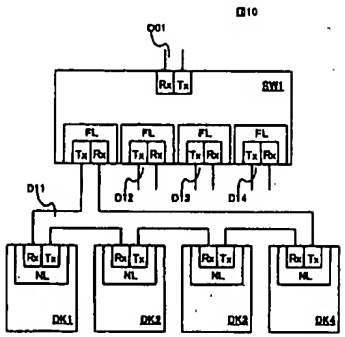
【図８】



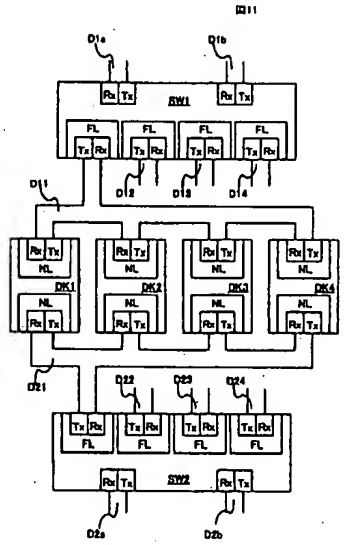
【図９】



【図１０】



【図１１】



【図１２】